[0225] Upon receiving data from the host system, at block 604, the processing logic reads a portion of the data received from the host system. As discussed above, the portion of the data may serve as an indicator for one or more properties associated with the underlying data received from the host system. In some embodiments, the processing logic may generate a histogram for the data received from the host system, as previously described at FIG. 5.

[0226] At block 606, the processing logic identifies a property associated with the data based on reading the portion of the data. In embodiments, the processing logic may identify the property based on the numerical representation of the generated histogram associated with the data.

[0227] At block 608, the processing logic allocates the data to a data segment based on the identified property. The processing logic may allocate the data to a data segment that includes other data allocated to the data segment that has one or more similar properties. For example, the processing logic may allocate data including mostly database tables to a data segment that includes other data that includes mostly database tables. In another example, the processing logic may allocate uncompressible data (e.g., images, videos, etc.) to a data segment that includes other data that is uncompressible.

[0228] At block 610, the processing logic stores the data segment that includes the allocated data at a storage device of a storage system.

[0229] While this specification contains many specific implementation details, these should not be construed as limitations on the scope of any inventions or of what may be claimed, but rather as descriptions of features specific to particular embodiments of particular inventions. Certain features that are described in this specification in the context of separate embodiments can also be implemented in combination in a single embodiment. Conversely, various features that are described in the context of a single embodiment can also be implemented in multiple embodiments separately or in any suitable subcombination. Moreover, although features may be described above as acting in certain combinations and even initially claimed as such, one or more features from a claimed combination can in some cases be excised from the combination, and the claimed combination may be directed to a subcombination or variation of a subcombination.

[0230] Similarly, while operations are depicted in the drawings in a particular order, this should not be understood as requiring that such operations be performed in the particular order shown or in sequential order, or that all illustrated operations be performed, to achieve desirable results. In certain circumstances, multitasking and parallel processing may be advantageous. Moreover, the separation of various system components in the embodiments described above should not be understood as requiring such separation in all embodiments, and it should be understood that the described program components and systems can generally be integrated together in a single software product or packaged into multiple software products.

[0231] Thus, particular embodiments of the subject matter have been described. Other embodiments are within the scope of the following claims. In some cases, the actions recited in the claims can be performed in a different order and still achieve desirable results. In addition, the processes depicted in the accompanying figures do not necessarily require the particular order shown, or sequential order, to achieve desirable results. In certain implementations, multitasking and parallel processing may be advantageous.

What is claimed is:

1. A storage system comprising:
a plurality of solid-state storage devices; and
a storage controller operatively coupled to the plurality of solid-state storage devices, the storage controller comprising a processing device, the processing device to:
identify a plurality of data blocks stored at one or more solid-state storage devices of the plurality of solid state storage devices;
read a portion of data from each data block of the plurality of data blocks;
determine a corresponding property for each data block of the plurality of data blocks based on reading the portion of the data;
identify a set of data blocks from the plurality of data blocks, wherein each data block of the set of data blocks is associated with a first corresponding property; and
store the set of data blocks at a data segment.

2. The storage system of claim 1, wherein the processing device is further to:
perform a compression operation on the set of data blocks at the data segment.

3. The storage system of claim 1, wherein the processing device is further to:
initiate a garbage collection process, wherein identifying the plurality of data blocks stored at the one or more solid-state storage devices is in response to initiating the garbage collection process.

4. The storage system of claim 1, wherein to read the portion of the data from each data block, the processing device is further to:
read a plurality of top bits of data from each data block of the plurality of data blocks.

5. The storage system of claim 1, wherein the processing device is further to:
generate a histogram for each data block of the plurality of data blocks based on the corresponding property of each data block.

6. The storage system of claim 1, wherein the corresponding property for each data block of the plurality of data blocks corresponds to a language associated with the data stored at each data block of the plurality of data blocks.

7. The storage system of claim 1, wherein the processing device is further to:
identify a second set of data blocks from the plurality of data blocks, wherein each data block of the second set of data blocks is associated with a second property; and
store the second set of data blocks at a second data segment.

8. A method comprising:
identifying a plurality of data blocks stored at one or more solid-state storage devices;
reading a portion of data from each data block of the plurality of data blocks;
determining, by a processing device, a corresponding property for each data block of the plurality of data blocks based on reading the portion of the data;